

# ACCELERATED DENSITY MATRIX EXPANSIONS FOR BORN–OPPENHEIMER MOLECULAR DYNAMICS

EMANUEL H. RUBENSSON AND ANDERS M. N. NIKLASSON

**ABSTRACT.** An accelerated polynomial expansion scheme to construct the density matrix in quantum mechanical molecular dynamics simulations is proposed. The scheme is based on recursive density matrix expansions, e.g. [Phys. Rev. B. 66 (2002), p. 155115], which are accelerated by a scale-and-fold technique [J. Chem. Theory Comput. 7 (2011), p. 1233]. The acceleration scheme requires interior eigenvalue estimates, which may be expensive and cumbersome to come by. Here we show how such eigenvalue estimates can be extracted from the recursive expansion by a simple and robust procedure at a negligible computational cost. Our method is illustrated with density functional tight-binding Born–Oppenheimer molecular dynamics simulations, where the computational effort is dominated by the density matrix construction. In our analysis we identify two different phases of the recursive polynomial expansion, the conditioning and purification phases, and we show that the acceleration represents an improvement of the conditioning phase, which typically gives a significant reduction of the computational cost.

## 1. INTRODUCTION

With the fast growth of computational processing power, atomistic simulations based on quantum mechanical calculations of the electronic structure have become a powerful approach to the study of a broad range of problems in materials science, chemistry, and biology [26, 33, 64]. Nevertheless, the computational cost associated with electronic structure calculations normally limits applications to fairly small systems. In particular, the cubic,  $O(N^3)$ , scaling of the computational cost as a function of the number of atoms,  $N$ , for the regular solution of the quantum mechanical eigenvalue problem is considered to be a most limiting factor. A number of different electronic structure technologies have therefore been developed that circumvent this bottleneck with a computational effort that scales only linearly with the system size [6, 18]. The reduction in the cost is typically achieved by utilizing sparse matrix algebra in an iterative construction of the density matrix, which avoids the full regular solution of the quantum mechanical eigenvalue problem. The matrix sparsity arise in localized atomic basis set representations due to the short-range character of the electronic wavefunctions for non-metallic materials [3, 27, 28]. With linear scaling techniques it is today possible to perform electronic structure calculations involving millions of atoms [5, 65].

An ideal linear scaling method should combine several important properties: *i*) it should be well tailored for high-performance on heterogeneous multi-core architectures, *ii*) require little intermediate memory storage, *iii*) allow well controlled numerical accuracy, and *iv*) have a low computational prefactor. Among several

linear scaling techniques [2, 4, 10, 11, 22, 24, 25, 31, 34, 45], the recursive second-order spectral projection method (SP2) [37], which sometimes also is referred to as the trace-correcting purification algorithm, represents a surprisingly simple and efficient choice [57]. The SP2 algorithm allows for control of errors arising from an approximate sparse matrix algebra [41, 54, 55] and it has a low computational prefactor. Even in the dense (non-sparse) limit the SP2 algorithm outperforms regular lapack diagonalization techniques on multi-core platforms, both with respect to speed and accuracy [9]. The computational cost of the SP2 algorithm is dominated by a sequence of matrix-matrix multiplications; typically between 20 to 40 multiplications are needed. Highly optimized linear algebra subroutines can perform dense matrix-matrix multiplications with close to peak performance on multi-core platforms. Thus, even without utilizing linear scaling sparse matrix algebra, the SP2 algorithm has excellent performance and is well suited for modern computational architectures [9].

Recently Rubensson proposed an acceleration technique that can further boost the performance of the SP2 algorithm, as well as other recursive polynomial expansion schemes [52]. The acceleration is based on the idea to allow for nonmonotonicity in the interval of interest, which gives additional flexibility in the choice of recursive expansion polynomials. By the use of a scale and fold technique, i.e. by shifting and re-scaling the SP2 projection polynomials, the number of matrix-matrix multiplications required by the SP2 algorithm is significantly reduced. This boosting technique requires estimates of interior eigenvalues that may be expensive and cumbersome to come by. In this article we show how such eigenvalue estimates can be extracted by a simple procedure from the recursive polynomial expansion at negligible computational cost.

Our scheme is particularly useful in molecular dynamics simulations. A molecular dynamics simulation often involves hundreds of thousands of time steps and therefore requires a low computational cost for the calculation of each new configuration. This is provided by the accelerated SP2 algorithm where eigenvalue estimates that are necessary for the acceleration can be extracted from previous time steps at practically no extra cost. The scheme is illustrated with density functional tight-binding molecular dynamics simulations, where the computational effort is dominated by the construction of the electronic density matrix through the recursive polynomial expansion. All the simulations are performed using dense matrix algebra in the computational framework of extended Lagrangian Born–Oppenheimer molecular dynamics [38, 61], which provides accurate and stable molecular trajectories with long-term energy conservation.

The article is outlined as follows: first we describe the problem and the recursive SP2 expansion of the density matrix that is used iteratively in the self-consistent calculation of the electronic ground state. Thereafter the acceleration technique is presented and we show how the required interior eigenvalues can be estimated from the recursion. The article goes on with a presentation of the extended Lagrangian formulation of Born–Oppenheimer molecular dynamics and we illustrate the behavior of our scheme in some quantum mechanical molecular dynamics simulations before we end with a discussion and some concluding remarks.

## 2. THE DENSITY MATRIX

The single-particle density matrix at zero electronic temperature can be defined as a matrix function of the effective Hamiltonian matrix,

$$(1) \quad D = \theta(\mu I - H),$$

where  $\theta(x)$  is the Heaviside step function and  $\mu$  is the chemical potential or Fermi level. We restrict our applications to non-metallic materials and assume that  $H$  does not have degenerate eigenvalues at  $\mu$ . We further assume that all matrices are Hermitian. We will refer to the smallest eigenvalue above  $\mu$  as the lowest unoccupied molecular orbital (lumo) eigenvalue and to the largest eigenvalue below  $\mu$  as the highest occupied molecular orbital (homo) eigenvalue. The gap between the homo and lumo eigenvalues is referred to as the homo-lumo gap. See Figure 1 for an illustration.

The density matrix is the matrix for orthogonal projection onto the occupied subspace spanned by the eigenvectors of  $H$  that correspond to eigenvalues smaller than  $\mu$ . The dimension of the occupied subspace is equal to the number of occupied electron orbitals,  $N_{\text{occ}}$ . Usually, the number of electrons is given as input to the program which has to automatically adjust the chemical potential so that a correct occupation number is achieved.

In general, the method of choice for computation of matrix functions of symmetric matrices is diagonalization [20, p. 84]. In short, an eigendecomposition  $H = V\Lambda V^T$  is computed (with  $\Lambda$  diagonal) and the matrix function can then be computed as  $f(H) = Vf(\Lambda)V^T$ . This is the standard method to solve (1) in the context of electronic structure calculations.

Since  $D$  is a projection matrix, an eigendecomposition, which possesses information about all eigenpairs of  $H$ , is not really needed to form  $D$ . Any basis of the occupied subspace could be used to construct  $D$ . Important for the accuracy however is to resolve the step at the chemical potential. In other words, while rotations of the eigenvectors *within* the occupied (or unoccupied) subspace do not affect the result at all, rotations corresponding to a leakage between the occupied and unoccupied subspaces do affect the accuracy, and are more likely to occur between eigenvectors corresponding to eigenvalues around the chemical potential. One would therefore expect the problem to become more difficult when the homo-lumo gap decreases. This is reflected by the condition number for the problem.

An appropriate condition number for the problem of computing the density matrix can be defined as

$$(2) \quad \kappa = \lim_{h \rightarrow 0} \sup_{A: \|A\| = \Delta\epsilon} \frac{\|\theta(\mu I - (H + hA)) - \theta(\mu I - H)\|}{h},$$

which can be evaluated to  $\kappa = \Delta\epsilon/\xi$ , where  $\Delta\epsilon = \lambda_{\text{max}} - \lambda_{\text{min}}$  is the spectral width and  $\xi$  is the homo-lumo gap of  $H$  [53, 55]. In the assessment of density matrix methods, it is therefore important to consider how the computational cost scales with the homo-lumo gap.

**2.1. The self-consistent field procedure.** In electronic structure methods such as Hartree-Fock [36, 51] or density functional theory [21, 29], the effective single-particle Hamiltonian matrix depends on the electronic configuration, which is determined by the density matrix. If the density matrix used to construct the Hamiltonian matrix is equal to the density matrix given by (1), we say that the solution

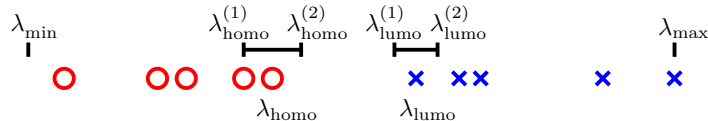


FIGURE 1. Eigenvalues of the effective Hamiltonian matrix  $H$  corresponding to occupied (red circles) and unoccupied (blue crosses) orbitals.  $(\lambda_{\text{homo}}^{(1)}, \lambda_{\text{homo}}^{(2)})$  and  $(\lambda_{\text{lumo}}^{(1)}, \lambda_{\text{lumo}}^{(2)})$  are intervals containing the homo and lumo eigenvalues, respectively. The chemical potential  $\mu$  is a number between  $\lambda_{\text{homo}}$  and  $\lambda_{\text{lumo}}$ .  $\lambda_{\text{min}}$  and  $\lambda_{\text{max}}$  are either the extremal eigenvalues or lower and upper bounds thereof. We will refer to  $\lambda_{\text{homo}}^{(2)}$  and  $\lambda_{\text{lumo}}^{(1)}$  as *inner bounds*, and  $\lambda_{\text{homo}}^{(1)}$  and  $\lambda_{\text{lumo}}^{(2)}$  as *outer bounds*.

is self-consistent. Finding the correct self-consistent electronic ground state therefore requires an iterative procedure, with repeated computation of Hamiltonian and density matrices,

$$(3) \quad D_0 \rightarrow H_1 \rightarrow D_1 \rightarrow H_2 \rightarrow D_2 \rightarrow \dots$$

In its simplest form, this self-consistent field procedure is a fixed-point iteration, where in each iteration the most recent density or Hamiltonian matrix is used as input to the next step. Usually, however, the Hamiltonian matrix is in each step taken as a linear combination of the most recent and previous Hamiltonian matrices to reduce the number of iterations needed to reach convergence. Popular mixing schemes include simple linear mixing, the direct inversion in the iterative subspace (DIIS) method [49], and Broyden mixing [7, 23].

In Born–Oppenheimer molecular dynamics simulations [33], the self-consistent field procedure is employed in each time step, as the electronic ground state solution is needed in the computation of the interatomic forces. In the context of the present work it does not matter if the sequence of matrices in (3) is generated by a self-consistent field optimization or if it is coming from molecular dynamics time stepping, as long as two successive Hamiltonians are reasonably similar to each other. The important point is that there is some procedure generating a sequence of Hamiltonian matrices  $H_1, H_2, \dots$  for which the corresponding density matrices need to be computed via (1).

### 3. RECURSIVE POLYNOMIAL EXPANSION OF THE DENSITY MATRIX

There are several alternatives to the eigenvalue decomposition for the calculation of the matrix step function in (1). In the context of linear scaling electronic structure theory some of the first techniques were based on serial Chebyshev expansions [18, 19, 59, 60], where the sparsity of the Hamiltonian can be used efficiently to achieve linear scaling complexity in the calculation of the Chebyshev polynomials through their recurrence relation. The degree of the Chebyshev polynomial required to reach a certain accuracy is proportional to  $\kappa = \Delta\epsilon/\xi$  [19]. The computational

cost therefore scales as  $\mathcal{O}(\kappa)$  or at best  $\mathcal{O}(\sqrt{\kappa})$  if the matrix polynomial is explicitly evaluated with the Paterson–Stockmeyer method [32, 48].

A difficulty with the Chebyshev techniques is that the chemical potential in (1) needs to be known in advance or has to be adjusted iteratively in order to achieve the correct occupation number. Palser and Manolopoulos proposed a scheme [46] that overcomes this problem through a recursive expansion based on the third-order purification polynomial  $3X^2 - 2X^3$  that originally was designed by McWeeny to adjust the fractional occupation of approximate density matrices to represent pure ensembles [35, 36]. In the Palser–Manolopoulos (PM) scheme (as well as in the original purification method by McWeeny) the density matrix is given through a rapid recursive expansion of the Heaviside step function where

$$(4) \quad \theta(\mu I - H) = \lim_{n \rightarrow \infty} f_n(f_{n-1}(\dots f_0(H) \dots)).$$

By adjusting  $X_0 = f_0(H)$  such that  $X_0$  has the correct trace, i.e. the desired occupation,  $N_{\text{occ}}$ , the successive McWeeny polynomials,  $f_n(X_{n-1})$ , can be modified to preserve the trace. In this trace conserving (or canonical) “purification” scheme the step is automatically formed at the chemical potential and no prior knowledge of the chemical potential is required. By using sparse matrix algebra each polynomial recursion  $X_n = f_n(X_{n-1})$  can be calculated with linear scaling effort.

A problem with the Palser–Manolopoulos scheme is that it is slow at low and high occupation, i. e. when  $N_{\text{occ}}/N$  is close to 0 or 1. A solution to this problem was offered by the recursive SP2 expansion scheme by Niklasson [37]. Instead of using third (or higher) order polynomials that simultaneously project the occupied eigenvalue spectrum of  $H$  towards 1 and the unoccupied towards 0, the SP2 scheme uses a combination of two second order spectral projection polynomials,

$$(5) \quad f_n(X) = \begin{cases} X^2 \\ 2X - X^2 \end{cases}$$

that either projects the eigenvalues towards 1 for  $f_n(X) = 2X - X^2$  or towards 0 for  $f_n(X) = X^2$ , see Figure 2. It is easy to see that any recursive combination of these two polynomials converges to a step function with the step formed in the interval  $[0, 1]$ . Moreover, since the trace of  $X^2$  is always smaller than the trace of  $X$ , given that all eigenvalues of  $X$  are in  $[0, 1]$ , and vice versa for  $2X - X^2$ , we can choose between the two polynomials to iteratively correct the occupation such that the trace of the converged expansion automatically has the desired value. The occupation correcting SP2 algorithm is described in some detail in Alg. 1. After an initial normalization, where  $\lambda_{\text{max/min}}$  are upper and lower bounds of the eigenspectrum of  $H$ , all eigenvalues of  $X$  are in reverse order in the interval  $[0, 1]$ . Thereafter we apply the spectral projection polynomials that are chosen to achieve the correct occupation at convergence. Prior knowledge of the chemical potential or the homo-lumo gap is not required.

Linear scaling computational complexity in density matrix methods is usually achieved by neglecting matrix elements deemed not to contribute significantly to the overall accuracy [54]. For recursive polynomial expansions on the form (4) the removal of matrix elements can be done in such a way that given a tolerance  $\varepsilon$ , an accuracy  $\|D - \tilde{D}\| \leq \varepsilon$  is guaranteed [54, 55]. Here,  $D$  and  $\tilde{D}$  are the exact and approximate density matrices respectively, and  $\|\cdot\|$  is a unitary invariant norm.

The polynomial expansion order increases very rapidly in the SP2 recursion. After only 30 matrix-matrix multiplications the expansion order is over 1 billion,

and the number of required multiplications increases only with the logarithm of the condition number, i. e. as  $\mathcal{O}(\log \kappa)$  [37]. Furthermore, the SP2 method performs well wherever the chemical potential may be located. In fact, the SP2 method is at its best at low and high occupation numbers. Also, the SP2 method is very memory-efficient as it requires a total of only two symmetric matrices in memory.

---

**Algorithm 1** Trace-correcting 2nd-order spectral projection expansion (SP2)

---

```

1:  $X = \frac{\lambda_{\max} I - H}{\lambda_{\max} - \lambda_{\min}}$ 
2: for  $i = 1, 2, \dots, n$  do
3:   if  $|\text{Tr}(X^2) - N_{\text{occ}}| < |\text{Tr}(2X - X^2) - N_{\text{occ}}|$  then
4:      $X = X^2$ 
5:   else
6:      $X = 2X - X^2$ 
7:   end if
8: end for
9: return  $X$ 

```

---

#### 4. ACCELERATION

The recursive expansion functions in (4) are usually taken as polynomials that have fixed points at 0 and 1, vanishing derivatives at 0 and/or 1, and that are monotonically increasing in the  $[0, 1]$  interval. Such polynomials are very good for bringing a near-idempotent matrix closer to idempotency. For the SP2 polynomials, for example, the asymptotic convergence rate is quadratic. However, the initial matrix is typically far from being idempotent, and most of the work in the recursive expansion is therefore spent on bringing the matrix near idempotency. For this task polynomials on the form described above are not optimal. By allowing for non-monotonicity in the projection polynomials the convergence can be boosted [52].

The basic idea behind the acceleration technique of the recursive polynomial expansion scheme is that prior knowledge of the eigenvalue distribution allows for a more optimized design of the projection polynomials that boosts convergence [52]. What is needed is, more precisely, estimates of the homo and lumo eigenvalues  $\lambda_{\text{homo}}$  and  $\lambda_{\text{lumo}}$  in Figure 1. If such estimates are available, the eigenspectrum can in each step be shifted and scaled so that the projection polynomials ( $X^2$  or  $2X - X^2$ ) fold the eigenspectrum, which results in a more rapid convergence. Effectively, the SP2 projection polynomials are shifted and rescaled as illustrated in Figure 2. The accelerated trace-correcting expansion is given by Alg. 2.

Too aggressive scaling can lead to the occupied and unoccupied eigenstates being mixed up. Therefore, a lower bound of the homo eigenvalue and an upper bound of the lumo eigenvalue,  $\lambda_{\text{homo}}^{(1)}$  and  $\lambda_{\text{lumo}}^{(2)}$  in Figure 1, are used in the accelerated algorithm. This makes the algorithm robust with respect to the homo-lumo estimates. If only loose bounds of the homo and lumo eigenvalues are available, the acceleration can still be used. As the algorithm is formulated, loose bounds will never lead to reduced efficiency or accuracy compared to the original SP2 algorithm, but tighter bounds result in a more efficient acceleration.

The process of bringing the matrix to idempotency can be seen as consisting of two different phases, that we shall refer to as the *conditioning phase* and the

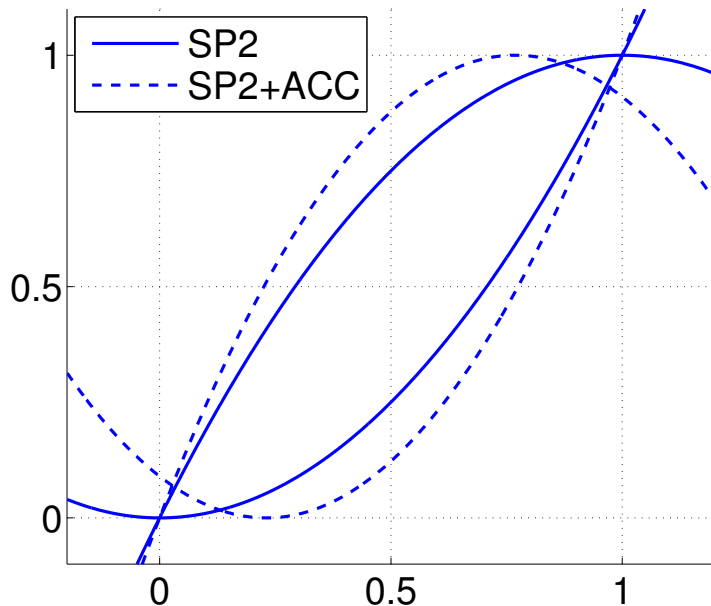


FIGURE 2. The regular SP2 projection polynomials  $X^2$  and  $2X - X^2$  (solid lines). The shifted and rescaled polynomials used in the accelerated recursive expansion,  $((1-\alpha)I + \alpha X)^2$  and  $2\alpha X - (\alpha X)^2$  (dashed lines). Here, the scaling parameter was taken as  $\alpha = 1.3$ . In practice, this parameter is adjusted during the course of the recursive expansion to give the best possible acceleration, see Alg. 2.

*purification phase.* In the first phase the idempotency error  $\|X - X^2\|$  is not reduced, but the condition number of the problem is lowered. Note that in each recursive expansion step, the problem that remains to be solved can be seen as an independent matrix step function problem with an associated condition number. The recursive expansion procedure may then be seen as a procedure to reduce the condition number to 1. When the condition number is close to 1, and the conditioning can no longer be substantially improved, we enter the purification phase in which the idempotency error is reduced. This behavior is illustrated in Figure 3.

It is clear that the acceleration represents an improvement of the conditioning phase which can be explained by a steeper slope of the projection polynomials at the chemical potential [37], where the occupied states are separated from the unoccupied states. As the condition number comes close to 1, and we enter the purification phase, the acceleration parameter  $\alpha_i$  in Alg. 2 comes close to 1 and the method becomes essentially equivalent to the original method without acceleration. Thus, the acceleration is automatically turned off when the purification phase is reached and the quadratic convergence of the SP2 method sets in.

**Algorithm 2** Accelerated expansion (SP2+ACC)

---

```

1:  $X = \frac{\lambda_{\max} I - H}{\lambda_{\max} - \lambda_{\min}}$ 
2:  $x_1 = \frac{\lambda_{\max} - \lambda_{\text{homo}}^{(1)}}{\lambda_{\max} - \lambda_{\min}}, \quad x_2 = \frac{\lambda_{\max} - \lambda_{\text{lumo}}^{(2)}}{\lambda_{\max} - \lambda_{\min}}$ 
3: for  $i = 1, 2, \dots, n$  do
4:   if  $|\text{Tr}(X^2) - N_{\text{occ}}| < |\text{Tr}(2X - X^2) - N_{\text{occ}}|$  then
5:      $p_i = 1$ 
6:      $\alpha_i = 2/(2 - x_2)$ 
7:      $X = (1 - \alpha_i)I + \alpha_i X$  (shift and scale)
8:      $X = X^2$ 
9:      $x_k = (1 - \alpha_i) + \alpha_i x_k, \quad k = 1, 2$ 
10:   else
11:      $p_i = 0$ 
12:      $\alpha_i = 2/(1 + x_1)$ 
13:      $X = \alpha_i X$  (scale)
14:      $X = 2X - X^2$ 
15:      $x_k = \alpha_i x_k, \quad k = 1, 2$ 
16:   end if
17:    $v_i = \|X - X^2\|_F$ 
18:    $w_i = \text{Tr}(X - X^2)$ 
19: end for
20: return  $X$ 

```

---

## 5. EIGENVALUE ESTIMATES

The drawback with the acceleration technique is that estimates of interior eigenvalues are needed. Standard iterative methods, such as e.g. the Lanczos method, provide rapid convergence to well separated extremal eigenvalues. When interior eigenvalues are to be computed, spectral transformations are typically employed to move the desired eigenvalues to the ends of the eigenspectrum. Transformations include shift-and-invert operators and various polynomial filters, see e.g. [1, 14]. The computational overhead that would be incurred to the recursive expansion by incorporating iterative eigenvalue methods depends on several factors such as matrix sparsity and hardware and software details. In many cases the computational overhead would be significant. However, in the context of recursive polynomial expansions one may take advantage of the formation of a step at the chemical potential. The recursive expansion leads to extremely good separation between eigenvalues around the chemical potential which can be utilized to drastically reduce the number of Lanczos iterations [56]. Nevertheless, even if it is possible to implement such schemes with good efficiency, the use of *any* iterative method for eigenvalue calculation would add to the complexity of the method.

Eigenvalue estimates can also be cheaply computed directly from the entries of the matrix, for example on the basis of Gershgorin's circle theorem. Unfortunately, Gershgorin's theorem rarely provides useful information about interior eigenvalues, at least not in the context of the present work.

In this section, we propose a new simple and efficient approach to calculate estimates of the homo and lumo eigenvalues when the recursive expansion is used in



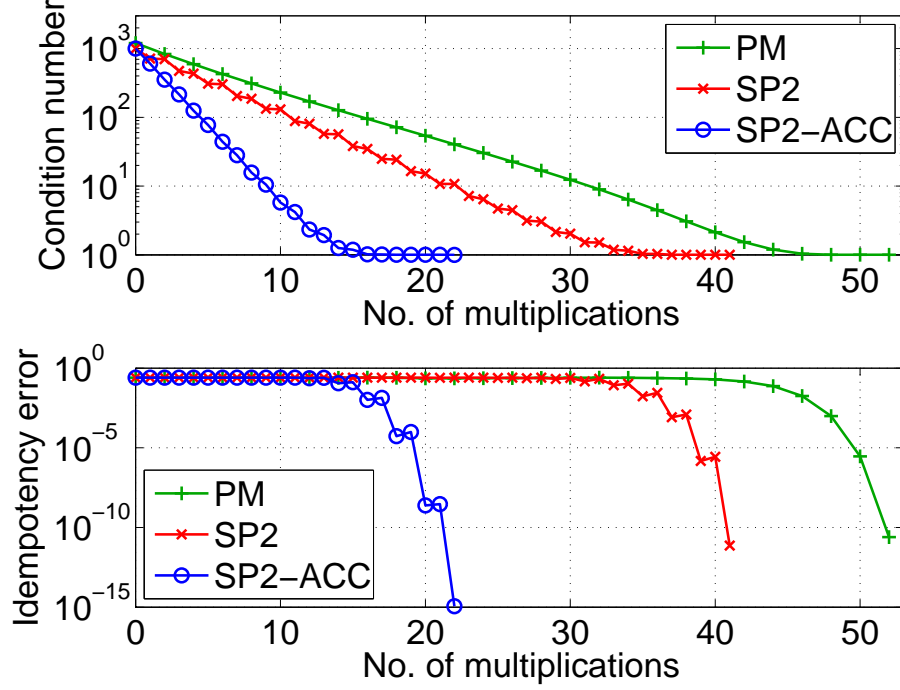


FIGURE 3. Illustration of the two phases of the recursive expansion. In the conditioning phase, the condition number  $\kappa = \Delta\epsilon/\xi$  is lowered but the idempotency error  $\|X - X^2\|_2$  is not reduced. When the condition number is close to 1 we have reached the purification phase, where the idempotency error decreases quadratically. The test calculations were performed for a test Hamiltonian for which the condition number  $\Delta\epsilon/\xi = 1000$  and with the chemical potential located at  $\mu = \lambda_{\min} + 0.3(\lambda_{\max} - \lambda_{\min})$ . The tested methods are the trace-conserving or canonical scheme (PM) [46], and the trace-correcting scheme with (SP2-ACC) and without (SP2) acceleration.

an outer loop such as for example molecular dynamics time stepping. The proposed method consists of two parts. The first is calculation of accurate estimates directly from matrix entries in the recursive expansion of the density matrix. The second is propagation of eigenvalue estimates between time steps (or SCF iterations).

### 5.1. Accurate interior eigenvalue estimates directly from matrix entries.

The proposed method is based on the following observations. We first recall that

$$(6) \quad \frac{\|A\|_F}{\sqrt{N}} \leq \|A\|_2 \leq \|A\|_F.$$

Now, let  $\{\lambda_j\}$  be the eigenvalues of  $X_i$ . Then,

$$(7) \quad \|X_i - X_i^2\|_F = \sqrt{\sum_j (\lambda_j - \lambda_j^2)^2}$$

and

$$(8) \quad \|X_i - X_i^2\|_2 = \max_j |\lambda_j - \lambda_j^2|.$$

The spectral norm in (8) is given by the eigenvalue closest to 0.5. In the last iterations of the density matrix recursion, this eigenvalue is either the homo or the lumo eigenvalue. Either the homo or the lumo eigenvalue can then be calculated via calculation of the norm in (8). In the last recursive expansion iterations, most of the eigenvalues are very close to zero and one, and the sum over all eigenvalues in (7) is dominated by the few eigenvalues furthest from convergence. This is illustrated in Figure 4. As we get closer to convergence, the Frobenius norm therefore becomes a better and better estimate of the spectral norm. The Frobenius norm of  $X_i - X_i^2$  in the final iterations can then be used to obtain estimates to the homo and lumo eigenvalues. Once the Frobenius norm has been computed, upper and lower bounds of  $\|X_i - X_i^2\|_2$ , given by (6), can be translated to bounds for the homo or lumo eigenvalues by taking the inverse of the recursive expansion [56]. This gives us a very sharp upper bound. However, for the lower bound in (6), we can do better.

Let  $\eta_j = \lambda_j - \lambda_j^2$ ,  $j = 1, 2, \dots, N$  be the eigenvalues of  $X_i - X_i^2$ . Since  $\eta_j \geq 0$ ,  $j = 1, 2, \dots, N$ , we have that

$$(9) \quad \sum_j \eta_j^2 \leq \max_j \eta_j \sum_k \eta_k$$

and since

$$(10) \quad \sum_j \eta_j^2 = \|X_i - X_i^2\|_F^2,$$

$$(11) \quad \sum_j \eta_j = \text{Tr}(X_i - X_i^2)$$

and

$$(12) \quad \|X_i - X_i^2\|_2 = \max_j \eta_j$$

we have that

$$(13) \quad \frac{\|X_i - X_i^2\|_F^2}{\text{Tr}(X_i - X_i^2)} \leq \|X_i - X_i^2\|_2.$$

Both the lower bound in (13) and the lower bound  $(\frac{\|X_i - X_i^2\|_F}{\sqrt{N}} \leq \|X_i - X_i^2\|_2)$  in (6) are sharp in the sense that if all eigenvalues of  $X_i - X_i^2$  are equal, then the left hand side equals the right hand side. However, the bound in (13) is sharp also in the sense that if  $X_i - X_i^2$  has only one nonzero eigenvalue, then the left hand side equals the right hand side as well. In the present context, many eigenvalues are expected to be zero, and we therefore expect the bound in (13) to be much better.

Before we can formulate an algorithm we need to address a few issues. The first question is which iterations that should be considered to be the last iterations, so that we can be sure that we compute the homo or the lumo eigenvalue and not some other eigenvalue.

From Alg. 2, we have  $v_i = \|X_i - X_i^2\|_F$ . By (6) we have that  $v_i \geq \|X_i - X_i^2\|_2 = \lambda - \lambda^2$  where  $\lambda$  is the eigenvalue of  $X_i$  that is closest to 0.5. Assume that  $v_i < 1/4$ ,

then we have that the interval

$$(14) \quad \left[ \frac{1}{2} - \sqrt{\frac{1}{4} - v_i}, \frac{1}{2} + \sqrt{\frac{1}{4} - v_i} \right]$$

is free from eigenvalues.

Assume now that in iteration  $i + 1$ ,  $\lambda_{\text{homo}} > 0.5$  and that  $\lambda_{\text{lumo}} < 0.5$ , and that the interval  $[\gamma, 1 - \gamma]$  is free from eigenvalues, where  $\gamma$  is a parameter between 0 and 0.5. If the interval  $[\gamma, 1 - \gamma]$  is free from eigenvalues also in iteration  $i$  and it is not possible for an eigenvalue to jump from a value larger than  $1 - \gamma$  to a value smaller than  $\gamma$  and vice versa in one iteration, then we know that  $\lambda_{\text{homo}} > 0.5$  and that  $\lambda_{\text{lumo}} < 0.5$  also in iteration  $i$ .

An eigenvalue can only jump from a value larger than  $1 - \gamma$  to a value smaller than  $\gamma$  and vice versa in one iteration provided that  $\gamma \geq 6 - 4\sqrt{2} \approx 0.343146$ . This value is given by solving

$$(15) \quad \alpha^2 \lambda^2 + 2\alpha(1 - \alpha)\lambda + (1 - \alpha)^2 = \gamma$$

where  $\lambda = 1 - \gamma$  and  $\alpha = \frac{2}{2 - \gamma}$ , for  $\gamma \in [0, 0.5]$ .

Thus, by the above reasoning and specifically by (14), the  $v_i$ -value can be used for eigenvalue estimation provided that

$$(16) \quad v_j < \gamma - \gamma^2, j = i, i + 1, \dots, n$$

with  $\gamma = 6 - 4\sqrt{2}$ . We only use the  $v_i$ -values for  $i$  that fulfills this criterion.

It would now be natural to ask the question how to know if it is the homo or lumo eigenvalue that dominates the Frobenius norm value in (7). Our response to this question is that we do not need to know. Regardless of whether the homo or lumo eigenvalue dominates the norm value,  $|\lambda_j - \lambda_j^2| \leq v_i$  holds for all eigenvalues. Since we know from above that  $\lambda_{\text{lumo}} < 0.5$  and that  $\lambda_{\text{homo}} > 0.5$ , we therefore have that  $\lambda_{\text{lumo}} \leq \frac{1}{2}(1 - \sqrt{1 - 4v_i})$  and that  $\lambda_{\text{homo}} \geq \frac{1}{2}(1 + \sqrt{1 - 4v_i})$ . Thus, the *inner bounds* for the homo and lumo eigenvalues always hold provided that the  $\gamma$ -criterion (16) is fulfilled.

Clearly, the same is not true for the *outer bounds*, see Figure 4. However, in each iteration, at least one of the outer bounds will be valid, either for the homo or the lumo eigenvalue. This means that provided that each of the homo and lumo eigenvalues dominate the norm value in at least one iteration for which the  $\gamma$ -criterion holds, the most conservative outer bound holds.

When the polynomial in each iteration is chosen based on the trace of the matrix, as in Algs. 1 and 2 it can in principle happen that one of the homo or lumo eigenvalues dominates the norm value in all the final iterations. This could happen if there are many eigenvalues clustered very close to either the homo or the lumo eigenvalue. This problem can be avoided by choosing polynomial based on the homo and lumo eigenvalue estimates, as in [55], which would also improve convergence.

Truncation and/or rounding errors cause eigenvalues to be disturbed. This has the effect that the eigenvalues at convergence will still deviate from their desired values of 0 and 1. Since the homo or the lumo eigenvalue dominates the norm value, this only matters in the very last iterations when the homo and lumo eigenvalues are very close to 1 and 0, respectively. As above this only leads to too conservative inner bounds and that the outer bounds may not hold in every iteration, which can be understood from Figure 4. In our algorithm we use the least conservative inner bounds and the most conservative outer bounds.

**5.2. Algorithm for homo and lumo eigenvalues.** The algorithm to calculate the estimates is described in Alg. 3. For each iteration  $i$  that the  $\gamma$ -criterion holds, candidates for upper and lower bounds of the homo and lumo eigenvalues of  $X_i$  are computed. For those values, the recursive expansion is carried out backwards so that potential bounds for the homo and lumo eigenvalues of  $X_{i-1}$ ,  $X_{i-2}$ ,  $\dots$ , and eventually  $X_0$  are obtained. Then, an attempt is made to improve the inner bounds to make them tighter and the outer bounds to make them more reliable, see the discussion above. Finally, when all potential bounds have been tested, the selected bounds are translated to bounds for the homo and lumo eigenvalues of  $H$ .

The algorithm requires some information that should be collected during the course of the recursive expansion:

- The parameters  $\lambda_{\min}$  and  $\lambda_{\max}$  used for the initial normalization.
- A boolean vector  $[p_1, p_2, \dots, p_n]$  specifying which polynomial that was used in each iteration.
- A vector  $[\alpha_1, \alpha_2, \dots, \alpha_n]$  containing the acceleration or scaling parameters used in each iteration.
- A vector  $[v_1, v_2, \dots, v_n]$  containing the Frobenius norms of  $X - X^2$  for each iteration.
- A vector  $[w_1, w_2, \dots, w_n]$  containing the traces of  $X - X^2$  for each iteration.

All this information can be extracted from the recursive expansion iterations at practically no extra cost, see Alg. 2.

**5.3. Propagation of eigenvalue estimates.** In the previous section, we developed a method to extract bounds for the homo and lumo eigenvalues from the recursive polynomial expansion. This means that the bounds are available only when the expansion has already completed. In order to accelerate the recursive expansion the bounds are needed in advance. However, as discussed in Section 2.1, the recursive expansions are often used repeatedly for a sequence of effective Hamiltonian matrices  $H_0, H_1, \dots$ . It is then possible to translate the eigenvalue bounds from one iteration to the next [55].

Let  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$  be the eigenvalues of  $H_{i-1}$  and  $\tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \dots \leq \tilde{\lambda}_N$  the eigenvalues of  $H_i$ . Then, by Weyl's theorem on eigenvalue movement [62],

$$(17) \quad |\lambda_j - \tilde{\lambda}_j| \leq \|H_i - H_{i-1}\|_2, \quad j = 1, 2, \dots, N.$$

Therefore, by expanding the intervals containing the homo and lumo eigenvalues of  $H_{i-1}$  with the norm  $\|H_i - H_{i-1}\|_2$ , we obtain intervals containing the homo and lumo eigenvalues of  $H_i$ . If the spectral norm cannot be readily obtained, the Frobenius norm can be used, see (6). If the norm is small (as is typically the case between SCF cycles), the acceleration of the following recursive expansion is hardly affected by this translation. If the norm is large (as is typically the case between time steps in a molecular dynamics simulation), the acceleration will not be optimal, but we have ensured the robustness of the expansion with respect to the homo and lumo estimates.

## 6. BORN–OPPENHEIMER MOLECULAR DYNAMICS

Among computationally feasible methods for molecular dynamics simulations, Born–Oppenheimer molecular dynamics based on self-consistent field methods, such as density functional theory [12, 21, 29, 47], is often considered a gold standard.

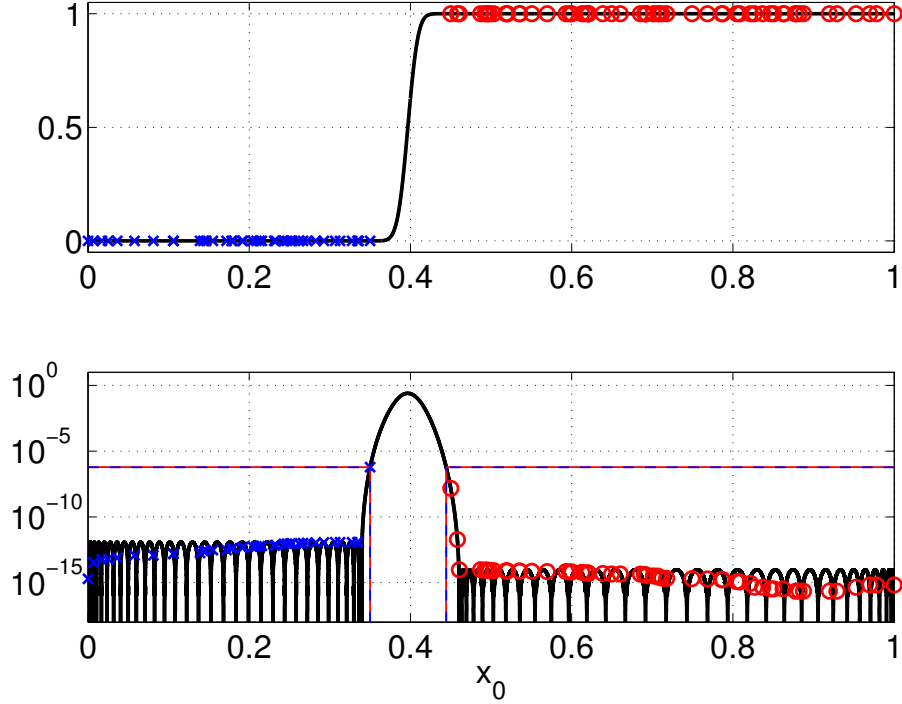


FIGURE 4. Mapping of eigenvalues by the accelerated expansion. Upper panel: Mapping  $f_{12}(f_{11}(\dots f_1(\lambda)\dots))$  of occupied (red circles) and unoccupied (blue crosses) eigenvalues  $\lambda$  of  $X_0$ . Lower panel: Absolute values of the eigenvalues of  $X_{12} - X_{12}^2$  as a function of the eigenvalues of  $X_0$ . The solid red lines and the dashed blue lines represent upper and lower bounds, respectively, of  $\|X_{12} - X_{12}^2\|_2$ , mapped back to the eigenspectrum of  $X_0$  to give information about the homo and lumo eigenvalues. The inner bounds for the homo and lumo eigenvalues, respectively, are both valid. In the twelfth iteration the outer bound for the lumo eigenvalue also holds. The outer bound for the homo eigenvalue, however, is not valid.

In Born–Oppenheimer molecular dynamics the forces acting on the atoms are explicitly quantum mechanical and calculated at the relaxed electronic ground state [33]. The electronic ground state is given through the self-consistent field optimization procedure, discussed in Section 2.1, which involves iterative solutions of the density matrix and Hamiltonian in (1). The self-consistent field optimization accounts for details in the charge distribution and is important for the accuracy of the interatomic forces.

**Algorithm 3** Eigenvalue computation

---

```

1:  $i = n$ 
2:  $x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 0$ 
3: while  $v_i < \gamma - \gamma^2$  do
4:    $y_1 = \frac{1}{2}(1 - \sqrt{1 - 4v_i^2/w_i})$ 
5:    $y_2 = \frac{1}{2}(1 - \sqrt{1 - 4v_i})$ 
6:    $y_3 = \frac{1}{2}(1 + \sqrt{1 - 4v_i})$ 
7:    $y_4 = \frac{1}{2}(1 + \sqrt{1 - 4v_i^2/w_i})$ 
8:   for  $j = i, i-1, \dots, 2, 1$  do
9:     if  $p_j$  then
10:        $y_k = \sqrt{y_k},$ 
11:        $y_k = (y_k - 1 + \alpha_j)/\alpha_j, \quad k = 1, 2, 3, 4$ 
12:     else
13:        $y_k = 1 - \sqrt{1 - y_k},$ 
14:        $y_k = y_k/\alpha_j, \quad k = 1, 2, 3, 4$ 
15:     end if
16:   end for
17:    $x_k = \min(x_k, y_k), \quad k = 1, 2$ 
18:    $x_k = \max(x_k, y_k), \quad k = 3, 4$ 
19:    $i = i - 1$ 
20: end while
21:  $\lambda_{\text{homo}}^{(1)} = \lambda_{\text{max}} - (\lambda_{\text{max}} - \lambda_{\text{min}})x_4$ 
22:  $\lambda_{\text{homo}}^{(2)} = \lambda_{\text{max}} - (\lambda_{\text{max}} - \lambda_{\text{min}})x_3$ 
23:  $\lambda_{\text{lumo}}^{(1)} = \lambda_{\text{max}} - (\lambda_{\text{max}} - \lambda_{\text{min}})x_2$ 
24:  $\lambda_{\text{lumo}}^{(2)} = \lambda_{\text{max}} - (\lambda_{\text{max}} - \lambda_{\text{min}})x_1$ 

```

---

In regular Born–Oppenheimer molecular dynamics the computational cost, which is dominated by the iterative self-consistent field optimization, is significantly reduced by using a good initial guess to the optimization procedure, which is naturally provided through an extrapolation of the ground state solutions from previous time steps. Since the iterative self-consistent field optimization in practice never is complete and always approximate, this extrapolation technique generates a time-dependent fictitious dynamics of the underlying electronic degrees of freedom. Because of the non-linearity of the self-consistent field optimization this process is irreversible, which typically is manifested in an unphysical systematic drift in the energy and phase space [42, 50], where the electronic degrees of freedom act like a heat sink or source, constantly removing or adding energy to the system. Only by increasing the number of self-consistent field iterations can the energy drift be reduced, though it never fully disappears. Recently an extended Lagrangian formulation of Born–Oppenheimer molecular dynamics was introduced that overcomes these fundamental problems by restoring time-reversibility and symplecticity to the dynamics [8, 38, 39, 40, 43, 61, 67].

Instead of using an extrapolation of the electronic ground state from previous time steps, as in regular Born–Oppenheimer molecular dynamics, an auxiliary electron density  $n(\mathbf{r})$  is introduced in the extended Lagrangian formulation as an extended dynamical variable that is centered around the ground state density  $\rho(\mathbf{r})$  through a harmonic well. The equations of motion for the nuclear coordinates

$\mathbf{R} = \{R_I\}$  and the extended density are given by

$$(18) \quad M_I \ddot{R}_I = -\frac{\partial U(\mathbf{R}; \rho)}{\partial R_I},$$

$$(19) \quad \ddot{n}(\mathbf{r}) = \omega^2 (\rho(\mathbf{r}) - n(\mathbf{r})),$$

where  $M_I$  are the nuclear masses and  $\omega$  is a frequency parameter that determines the curvature of the harmonic well.  $U(\mathbf{R}; \rho)$  is the potential energy term that describes the electronic and ionic energy at the self-consistent electronic ground state. The dots denote time-derivatives.

The extended Lagrangian molecular dynamics, (18) and (19), has several important features: *i)*  $n(\mathbf{r})$  moves in a harmonic well centered around the ground state density  $\rho(\mathbf{r})$  and is therefore a good initial guess to the iterative self-consistent field optimization procedure, *ii)* since  $n(\mathbf{r})$  is a dynamical variable, in contrast to  $\rho(\mathbf{r})$ , it can be integrated with a geometric reversible scheme in the same way as the nuclear coordinates [30, 38, 43], *iii)* in this way reversibility in the underlying propagation of the electronic degrees of freedom is not broken, and *iv)* since the equations of motion for the nuclear coordinate is identical to “exact” Born–Oppenheimer molecular dynamics, the total Born–Oppenheimer energy,

$$(20) \quad E_{\text{tot}} = \frac{1}{2} \sum_I M_I \dot{R}_I^2 + U(\mathbf{R}; \rho),$$

is still a constant of motion that should exhibit long-term stability [38].

Since our estimates of the homo-lumo eigenvalues, which are used for the scale-and-fold acceleration in the construction of the density matrix, are given from previous time steps, it is important to show that this additional artificial propagation does not have any effect on the reversibility and long-term energy conservation. Here we will demonstrate how the accelerated SP2 density matrix expansion using estimates of the homo-lumo eigenvalues from the previous time step is fully compatible with extended Lagrangian Born–Oppenheimer molecular dynamics, without causing any systematic drift in the energy. Our accelerated expansion algorithm therefore provides a significant reduction of the computational cost, while keeping the accuracy, for an already highly efficient scheme.

## 7. NUMERICAL EXPERIMENTS

In our numerical experiments the interaction potential  $U(\mathbf{R}; \rho)$  in (18) is based on the self-consistent density functional tight-binding (SCC-DFTB) formulation of density functional theory [13, 15, 16, 17, 66] as implemented in the program LATTE [8, 9, 39, 44, 58]. The equations of motion are integrated with the Verlet scheme, in its symplectic velocity formulation for the nuclear degrees of freedom and using a regular Verlet scheme for the electronic degrees of freedom with a dissipative force term that removes any accumulation of numerical noise [40, 61]. In all cases the SP2 algorithm, with (Alg. 2) or without (Alg. 1) acceleration, was used with the recursions terminated when the occupation change, as measured by the difference between using  $X^2$  or  $2X - X^2$ , was less than  $10^{-10}$ . Two self-consistent field iterations were used in each time step, which together with the force calculation sums up to totally three density matrix constructions per time step. The homo and lumo eigenvalue bounds were translated from one time step or self-consistent field iteration to the next as described in Section 5.3, using the Frobenius matrix norm

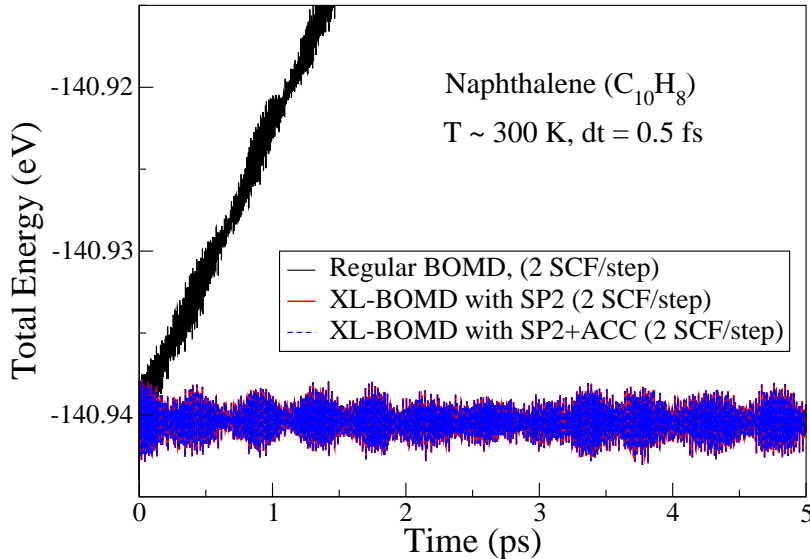


FIGURE 5. The conservation of the total energy, (20), for regular Born–Oppenheimer molecular dynamics (BOMD) and extended Lagrangian Born–Oppenheimer molecular dynamics (XL-BOMD), with and without the acceleration scheme (ACC) for the SP2 algorithm. The two lower curves (red and blue) are virtually identical.

in (17). All calculations were performed using regular dense (non-sparse) matrix algebra.

Figure 5 shows the stability of the total Born–Oppenheimer energy,  $E_{\text{tot}}$  in (20), either with regular Born–Oppenheimer molecular dynamics using the electronic ground state solution from the previous time step as an initial guess to the self-consistent field optimization or with extended Lagrangian Born–Oppenheimer molecular dynamics, (18) and (19), with or without the acceleration scheme. The systematic drift occurring in regular Born–Oppenheimer molecular dynamics vanishes in the extended Lagrangian formulation. Moreover, we find virtually no effect in the total energy from using the accelerated SP2 scheme with estimates of the homo and luma eigenvalues obtained from the previous iteration. The two graphs are practically identical even after thousands of time steps.

The gain in efficiency using the acceleration technique is significant. Figure 6 shows the average number of matrix-matrix multiplications required in the density matrix expansion in molecular dynamics simulations of various systems. For the simulation of the polyethylene molecule we chose a 100 carbon atom chain ( $C_{100}H_{202}$ ) and the methane example consisted of liquid methane ( $CH_4$ )<sub>100</sub>. All the simulations were performed with a time step of 0.5 fs using a structurally optimized ground state



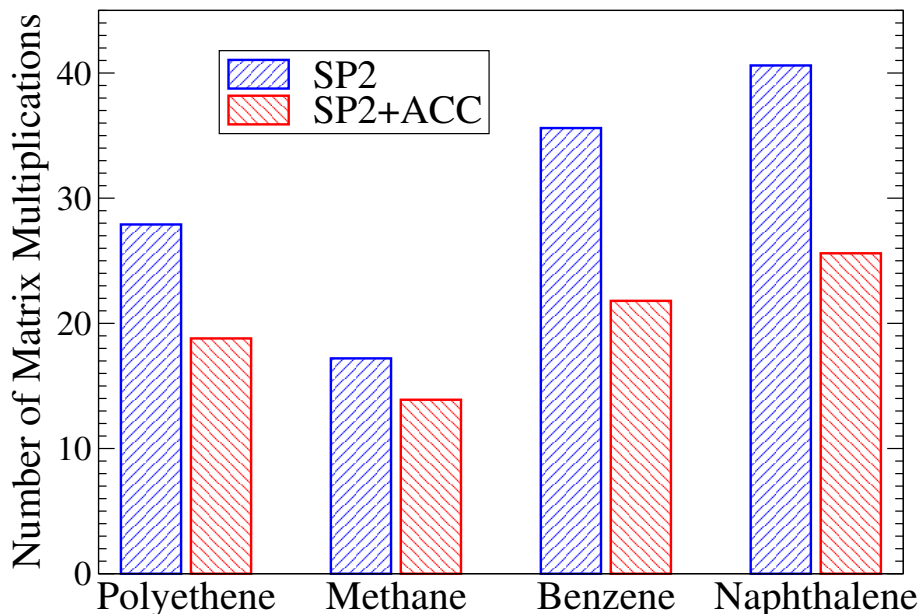


FIGURE 6. The average number of matrix-matrix multiplications of the SP2 algorithm during molecular dynamics simulations with and without acceleration.

configuration with an initial Gaussian velocity distribution corresponding to 550 K. The efficiency gain is governed by the size of the homo-lumo gap [52]. For small gap systems, like naphthalene, the scale-and-fold acceleration gives a significantly higher reduction of the cost compared to large gap systems, like methane. In other words, the acceleration works best for ill-conditioned systems, as defined in Section 2. This also means that there is a much smaller variation in the number of matrix multiplications required for the accelerated scheme. The work load is therefore not only smaller but also less sensitive to the particular choice of materials system. This also means that the computational cost is more evenly distributed over time in a molecular dynamics simulation when there are large variations in the homo-lumo gap.

Figure 7 (panel a-d) illustrates the effect of the acceleration as a function of simulation time for a system of liquid isocyanic acid. As seen in the two upper panels (a and b) the energetics between the regular and the accelerated scheme, as measured by the total energy or the temperature, are indistinguishable. Figure 7 (panel d) also shows how the number of matrix multiplications for the density matrix construction (that is required in the last self-consistent field iteration in each time step) varies over time. Variations in the work load may be sensitive to the choice of convergence criterion used to terminate the SP2 recursions, which here is set fairly tight. The accuracy of the homo-lumo gap estimate, shown in panel (c) is surprisingly good, in particular as measured by the inner (Inner)  $\lambda_{\text{homo}}^{(2)}$  and  $\lambda_{\text{lumo}}^{(1)}$  eigenvalues. The upper bound of the gap, given by the outer (Outer) eigenvalue bounds, is good but not as tight. The reason is that for the outer bounds we choose the most conservative values to make sure that we get valid bounds. It is in principle

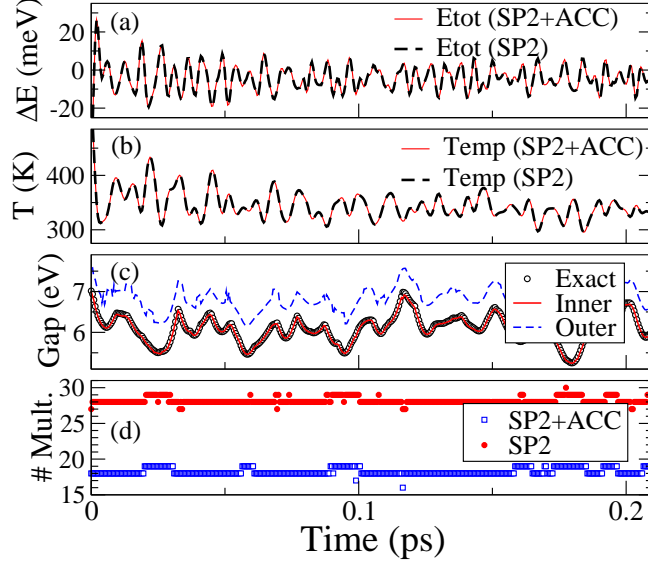


FIGURE 7. The fluctuations during a simulation run of liquid isocyanic acid ( $HCNO$ )<sub>56</sub> of the total energy (Etot in panel a), the temperature (Temp in panel b), the gap estimates (panel c), and the variation of the number of matrix-matrix multiplications (panel d) of the SP2 algorithm with and without acceleration (ACC).

possible to devise a scheme that keeps track of which eigenvalue (the homo or lumo eigenvalue) that correspond to the spectral norm of  $X - X^2$ . This information can be used to obtain more accurate outer homo-lumo bounds. However, such an algorithm would be more complicated with small gains in efficiency.

## 8. DISCUSSION

In this article we have devised a surprisingly simple technique to extract information about the homo-lumo gap that enables a cost-efficient practical application of the scale-and-fold acceleration technique in molecular dynamics simulations. The technique is quite general and should be applicable to a broad class of recursive density matrix expansion algorithms besides the SP2 scheme described here [52, 63]. Scale-and-fold acceleration is not the only way to improve or design efficient recursive expansion polynomials. It is in principle possible to construct a number of “optimal” recursive expansion schemes where the projection polynomials,  $\{f_n\}$ , in the recursive expansion (4) are defined, for example, by minimizing the error in some given norm for a certain number of recursion steps,  $N$ , and with a maximum number,  $M_{\max}$ , of necessary matrix multiplications,  $M$ , e.g.

$$(21) \quad \{f_n\}_{n=0}^N = \arg \min_{\{f_n\}} \{ \|\theta(\mu I - H) - f_N(f_{N-1}(\dots f_0(H)\dots))\|; M \leq M_{\max} \}.$$

With a detailed knowledge about the particular eigenvalue spectrum of  $H$  this problem can be solved using a number of various non-linear optimization techniques. The key issue with such an optimized recursion scheme, (21), is that we in general do not have prior knowledge of the interior eigenvalue distribution. Moreover, in a molecular dynamics simulation or a self-consistent field optimization the optimization would have to be performed on-the-fly in each iteration. Alternatively, optimizations for different classes of distributions could be made, e.g. for uniform eigenvalue distributions with various homo-lumo gaps and spectral bounds, which possibly could be parametrized. Our accelerated SP2 recursion scheme may very well belong to such a hypothetical class of optimized expansion methods.

In this article we have taken a cautious standpoint with respect to the homo and lumino estimates used in the acceleration. Although the inner bounds of the homo and lumino eigenvalues have proven to be extremely tight, we have used the more conservative outer bounds for the acceleration. Moreover, those outer bounds have been moved further away from the gap as a result of the propagation between time steps or self-consistent field iterations by (17). All those measures are not strictly necessary in order to avoid mixing up occupied and unoccupied states and lead to a less efficient acceleration. However, we also have to avoid folding extremal eigenvalues into the gap since this could lead to invalid homo and lumino eigenvalue bounds in the following iterations. With the conservative measures taken in this work this will never happen, though in practice it is often possible to use less restrictive homo-lumo estimates leading to a more efficient acceleration.

## 9. CONCLUSIONS

We have proposed a scheme to accelerate the construction of the density matrix in quantum mechanical molecular dynamics simulations. Our technique is based on the scale-and-fold acceleration method with estimates of the homo-lumo gap that are automatically acquired from the recursive expansion in the previous time step or self-consistent field iteration at a negligible computational cost. Our method was illustrated with density functional tight-binding molecular dynamics simulations, where the computational effort is dominated by the density matrix construction. Our scheme was found to be fully compatible with extended Lagrangian Born–Oppenheimer molecular dynamics and long-term energy stability. As a stand-alone technique, our method for calculating the homo-lumo gap estimates may also be useful in studies of materials properties.

The experiments presented in this work were all carried out using dense matrix algebra. The use of sparse matrix algebra to achieve a linear scaling algorithm involves additional issues regarding ways to bring about matrix sparsity. Recursive expansions with rigorous error control requires homo and lumino eigenvalue estimates [55]. Fortunately, this is the same information that is needed for the acceleration. Thus, the eigenvalue estimates of the present work will also be useful in schemes to select small matrix elements for removal to achieve error control in the linear scaling construction of the density matrix.

We identified two different phases in recursive density matrix expansions, the conditioning phase and the purification phase. With this point of view it was clear that the acceleration technique represents an improvement of the conditioning phase. In this way, the acceleration is analogous to preconditioners used with iterative solvers for linear systems. By using prior knowledge of the system, the

condition number of the problem is more rapidly reduced in the (pre-)conditioning phase, which gives a significant reduction of the computational cost, as demonstrated in quantum mechanical molecular dynamics simulations.

#### ACKNOWLEDGMENTS

Support from Göran Gustafsson's foundation, the Swedish research council (grant no. 621-2012-3861), Lisa and Carl-Gustav Esseens foundation, the Swedish national strategic e-science research program (eSSENCE), United States Department of Energy Office of Basic Energy Sciences, and the International Ten Bar Java group of T. Peery, as well as discussions with M. Cawkwell are gratefully acknowledged.

#### REFERENCES

- [1] Z. BAI, J. DEMMEL, J. DONGARRA, A. RUHE, AND H. VAN DER VORST, eds., *Templates for the solution of algebraic eigenvalue problems*, SIAM, Philadelphia, 2000.
- [2] M. BARRAULT, E. CANCÈS, W. W. HAGER, AND C. LE BRIS, *Multilevel domain decomposition for electronic structure calculations*, J. Comput. Phys., 222 (2007), pp. 86–109.
- [3] M. BENZI, P. BOITO, AND N. RAZOUK, *Decay properties of spectral projectors with applications to electronic structure*, SIAM Review, 55 (2013), pp. 3–64.
- [4] E. G. BIRGIN, J. M. MARTÍNEZ, L. MARTÍNEZ, AND G. B. ROCHA, *Sparse projected-gradient method as a linear-scaling low-memory alternative to diagonalization in self-consistent field electronic structure calculations*, J. Chem. Theory Comput., 9 (2013), pp. 1043–1051.
- [5] D. R. BOWLER AND T. MIYAZAKI, *Calculations for millions of atoms with density functional theory: linear scaling shows its potential*, J. Phys.: Condens. Matter, 22 (2010), p. 074207.
- [6] ———,  *$O(n)$  methods in electronic structure calculations*, Rep. Prog. Phys., 75 (2012), pp. 036503–036546.
- [7] C. G. BROYDEN, *A class of methods for solving nonlinear simultaneous equations*, Math. Comp., 19 (1965), p. 557.
- [8] M. J. CAWKWELL AND A. M. N. NIKLASSON, *Energy conserving, linear scaling Born–Oppenheimer molecular dynamics*, J. Chem. Phys., 137 (2012), p. 134105.
- [9] M. J. CAWKWELL, E. J. SANVILLE, S. M. MNISZEWSKI, AND A. M. N. NIKLASSON, *Computing the density matrix in electronic structure theory on graphics processing units*, J. Chem. Theory Comput., 8 (2012), p. 4094.
- [10] M. CHALLACOMBE, *A simplified density matrix minimization for linear scaling self-consistent field theory*, J. Chem. Phys., 110 (1999), pp. 2332–2342.
- [11] M. S. DAW, *Model for energetics of solids based on the density matrix*, Phys. Rev. B, 47 (1993), pp. 10895–10898.
- [12] R. M. DREIZLER AND K. U. GROSS, *Density-functional theory*, Springer Verlag, Berlin Heidelberg, 1990.
- [13] M. ELSTNER, D. PORESAG, G. JUNGNIKKEL, J. ELSTNER, M. HAUGK, T. FRAUENHEIM, S. SUHAI, AND G. SEIFERT, *Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties*, Phys. Rev. B, 58 (1998), p. 7260.
- [14] H. FANG AND Y. SAAD, *A filtered Lanczos procedure for extreme and interior eigenvalue problems*, SIAM J. Sci. Comput., 34 (2012), pp. A2220–A2246.
- [15] M. FINNIS, *Interatomic forces in condensed matter*, Oxford University Press, 2003.
- [16] M. FINNIS, A. T. PAXTON, M. METHFESSEL, AND M. VAN SCHILFGAARDE, *Crystal structures of zirconia from first principles and self-consistent tight binding*, Phys. Rev. Lett., 81 (1998), p. 5149.
- [17] T. FRAUENHEIM, G. SEIFERT, M. ELSTNER, Z. HAJNAL, G. JUNGNIKKEL, D. PORESAG, S. SUHAI, AND R. SCHOLZ, *A self-consistent charge density-functional based tight-binding method for predictive materials simulations in physics, chemistry and biology*, Phys. stat. sol. (b), 217 (2000), p. 41.
- [18] S. GOEDECKER, *Linear scaling electronic structure methods*, Rev. Mod. Phys., 71 (1999), pp. 1085–1123.
- [19] S. GOEDECKER AND L. COLOMBO, *Efficient linear scaling algorithm for tight-binding molecular-dynamics*, Phys. Rev. Lett., 73 (1994), pp. 122–125.

- [20] N. J. HIGHAM, *Functions of matrices: theory and computation*, SIAM, Philadelphia, 2008.
- [21] P. HOHENBERG AND W. KOHN, *Inhomogeneous electron gas*, Phys. Rev., 136 (1964), pp. B:864–B871.
- [22] A. HOLAS, *Transforms for idempotency purification of density matrices in linear-scaling electronic-structure calculations*, Chem. Phys. Lett., 340 (2001), pp. 552–558.
- [23] D. D. JOHNSON, *Modified Broyden’s method for accelerating convergence in self-consistent calculations*, Phys. Rev. B, 38 (1988), p. 12807.
- [24] J. KIM AND Y. JUNG, *Accelerated purification using generalized nonpurifying intermediate functions for large-scale self-consistent field calculations*, J. Chem. Theory Comput., 7 (2011), pp. 3853–3858.
- [25] J. KIM, F. MAURI, AND G. GALLI, *Total-energy global optimizations using nonorthogonal localized orbitals*, Phys. Rev. B, 52 (1995), pp. 1640–1648.
- [26] B. KIRCHNER, J. DI DIO PHILIPP, AND J. HUTTER, *Real-world predictions from ab initio molecular dynamics simulations*, Topocs. Curr. Chem., 307 (2012), p. 109.
- [27] W. KOHN, *Analytical properties of bloch waves and wannier functions*, Phys. Rev., 115 (1959), pp. 809–821.
- [28] ———, *Density functional and density matrix method scaling linearly with the number of atoms*, Phys. Rev. Lett., 76 (1996), p. 3168.
- [29] W. KOHN AND L. J. SHAM, *Self-consistent equations including exchange and correlation effects*, Phys. Rev., 140 (1965), p. 1133.
- [30] B. LEIMKUHLER AND S. REICH, *Simulating Hamiltonian Dynamics*, Cambridge University Press, 2004.
- [31] X.-P. LI, R. W. NUNES, AND D. VANDERBILT, *Density-matrix electronic-structure method with linear system-size scaling*, Phys. Rev. B, 47 (1993), pp. 10891–10894.
- [32] W. LIANG, C. SARAVANAN, Y. SHAO, R. BAER, A. T. BELL, AND M. HEAD-GORDON, *Improved Fermi operator expansion methods for fast electronic structure calculations*, J. Chem. Phys., 119 (2003), pp. 4117–4125.
- [33] D. MARX AND J. HUTTER, *Modern Methods and Algorithms of Quantum Chemistry*, ed. J. Grotendorst, John von Neumann Institute for Computing, Jülich, Germany, second ed., 2000.
- [34] D. A. MAZZIOTTI, *Towards idempotent reduced density matrices via particle-hole duality: Mcweeny’s purification and beyond*, Phys. Rev. E, 68 (2003), p. 066701.
- [35] R. MCWEENY, *The density matrix in self-consistent field theory. i. iterative construction of the density matrix*, Proc. R. Soc. London Ser. A, 235 (1956), pp. 496–509.
- [36] ———, *Some recent advances in density matrix theory*, Rev. Mod. Phys., 32 (1960), p. 335.
- [37] A. M. N. NIKLASSON, *Expansion algorithm for the density matrix*, Phys. Rev. B, 66 (2002), p. 155115.
- [38] A. M. N. NIKLASSON, *Extended Born–Oppenheimer molecular dynamics*, Phys. Rev. Lett., 100 (2008), p. 123004.
- [39] A. M. N. NIKLASSON AND M. J. CAWKWELL, *Fast method for quantum mechanical molecular dynamics*, Phys. Rev. B, 86 (2012), p. 174308.
- [40] A. M. N. NIKLASSON, P. STENETEG, A. ODELL, N. BOCK, M. CHALLACOMBE, C. J. TYMCZAK, E. HOLMSTRÖM, G. ZHENG, AND V. WEBER, *Time-reversible Born–Oppenheimer molecular dynamics with dissipation*, J. Chem. Phys., 130 (2009), p. 214109.
- [41] A. M. N. NIKLASSON, C. J. TYMCZAK, AND M. CHALLACOMBE, *Trace resetting density matrix purification in  $O(N)$  self-consistent-field theory*, J. Chem. Phys., 118 (2003), p. 8611.
- [42] A. M. N. NIKLASSON, C. J. TYMCZAK, AND M. CHALLACOMBE, *Time-reversible Born–Oppenheimer molecular dynamics*, Phys. Rev. Lett., 97 (2006), p. 123001.
- [43] A. ODELL, A. DELIN, B. JOHANSSON, N. BOCK, M. CHALLACOMBE, AND A. M. N. NIKLASSON, *Higher-order symplectic integration in Born–Oppenheimer molecular dynamics*, J. Chem. Phys., 131 (2009), p. 244106.
- [44] A. ODELL, A. DELIN, B. JOHANSSON, M. J. CAWKWELL, AND A. M. N. NIKLASSON, *Geometric integration in Born–Oppenheimer molecular dynamics*, 135 (2011), p. 224105. J. Chem. Phys.
- [45] P. ORDEJÓN, D. A. DRABOLD, M. P. GRUMBACH, AND R. M. MARTIN, *Unconstrained minimization approach for electronic computations that scales linearly with system size*, Phys. Rev. B, 48 (1993), pp. 14646–14649.
- [46] A. H. R. PALSER AND D. E. MANOLOPOULOS, *Canonical purification of the density matrix in electronic-structure theory*, Phys. Rev. B, 58 (1998), p. 12704.

- [47] R. G. PARR AND W. YANG, *Density-functional theory of atoms and molecules*, Oxford University Press, Oxford, 1989.
- [48] M. S. PATERSON AND L. STOCKMEYER, *On the number of nonscalar multiplications necessary to evaluate matrix polynomials*, SIAM J. Comput., 2 (1973), pp. 60–66.
- [49] P. PULAY, *Convergence acceleration of iterative sequences. the case of scf iteration*, Chem. Phys. Lett., 73 (1980), pp. 393–398.
- [50] P. PULAY AND G. FOGARASI, *Fock matrix dynamics*, 386 (2004), p. 272.
- [51] C. C. J. Roothaan, *New developments in molecular orbital theory*, Rev. Mod. Phys., 23 (1951), pp. 69–89.
- [52] E. H. RUBENSSON, *Nonmonotonic recursive polynomial expansions for linear scaling calculation of the density matrix*, J. Chem. Theory Comput., 7 (2011), pp. 1233–1236.
- [53] E. H. RUBENSSON, *Controlling errors in recursive Fermi–Dirac operator expansions with applications in electronic structure theory*, SIAM J. Sci. Comput., 34 (2012), pp. B1–B23.
- [54] E. H. RUBENSSON AND E. RUDBERG, *Bringing about matrix sparsity in linear scaling electronic structure calculations*, J. Comput. Chem., 32 (2011), pp. 1411–1423.
- [55] E. H. RUBENSSON, E. RUDBERG, AND P. SALEK, *Density matrix purification with rigorous error control*, J. Chem. Phys., 128 (2008), p. 074106.
- [56] E. H. RUBENSSON AND S. ZAHEDI, *Computation of interior eigenvalues in electronic structure calculations facilitated by density matrix purification*, J. Chem. Phys., 128 (2008), p. 176101.
- [57] E. RUDBERG AND E. H. RUBENSSON, *Assessment of density matrix methods for linear scaling electronic structure calculations*, J. Phys.: Condens. Matter, 23 (2011), p. 075502.
- [58] E. SANVILLE, N. BOCK, W. M. CHALLACOMBE, A. M. N. NIKLASSON, M. J. CAWKWELL, D. M. DATTELBAUM, AND S. SHEFFIELD, *Proceedings of the Fourteenth International Detonation Symposium*, Office of Naval Research, Arlington VA, ONR-351-10-185, 2010, pp. 91–101.
- [59] R. N. SILVER AND H. RÖDER, *Densities of states of mega-dimensional Hamiltonian matrices*, Int. J. Mod. Phys. C, 5 (1994), pp. 735–753.
- [60] R. N. SILVER, H. RÖDER, A. F. VOTER, AND J. D. KRESS, *Kernel polynomial approximations for densities of states and spectral functions*, Int. J. Comput. Phys., 124 (1996), p. 115.
- [61] P. STENETEG, I. A. ABRIKOSOV, V. WEBER, AND A. M. N. NIKLASSON, *Wave function extended lagrangian Born–Oppenheimer molecular dynamics*, Phys. Rev. B, 82 (2010), p. 075110.
- [62] G. W. STEWART AND J. SUN, *Matrix perturbation theory*, Academic Press, San Diego, 1990.
- [63] P. SURYANARAYANA, *Optimized purification for density matrix calculation*, Chem. Phys. Lett., 555 (2013), pp. 291–295.
- [64] M. TUCKERMAN, *Ab initio molecular dynamics: basic concepts, current trends and novel applications*, J. Phys.: Condens. Matter, 50 (2002), p. 1297.
- [65] J. VANDEVONDELE, U. BORSTNIK, AND J. HUTTER, *Linear scaling self-consistent field calculations with millions of atoms in the condensed phase*, J. Chem. Theory Comput., 8 (2012), pp. 3565–3573.
- [66] G. ZHENG, S. IRLE, AND K. MOROKUMA, *Performance of the DFTB method in comparison to DFT and semiempirical methods for geometries and energies of c20c86 fullerene isomers*, 412 (2005), p. 210.
- [67] G. ZHENG, A. M. N. NIKLASSON, AND M. KARPLUS, *Lagrangian formulation with dissipation of Born–Oppenheimer molecular dynamics using the density-functional tight-binding method*, J. Chem. Phys., 135 (2011), p. 044122.

*E-mail address*, E. H. Rubensson: [emanuel.rubensson@it.uu.se](mailto:emanuel.rubensson@it.uu.se)

(E. H. Rubensson) DIVISION OF SCIENTIFIC COMPUTING, DEPARTMENT OF INFORMATION TECHNOLOGY, UPPSALA UNIVERSITY, BOX 337, SE-751 05 UPPSALA, SWEDEN

*E-mail address*, A. M. N. Niklasson: [amn@lanl.gov](mailto:amn@lanl.gov)

(A. M. N. Niklasson) THEORETICAL DIVISION, LOS ALAMOS NATIONAL LABORATORY, LOS ALAMOS, NEW MEXICO 87545, USA